

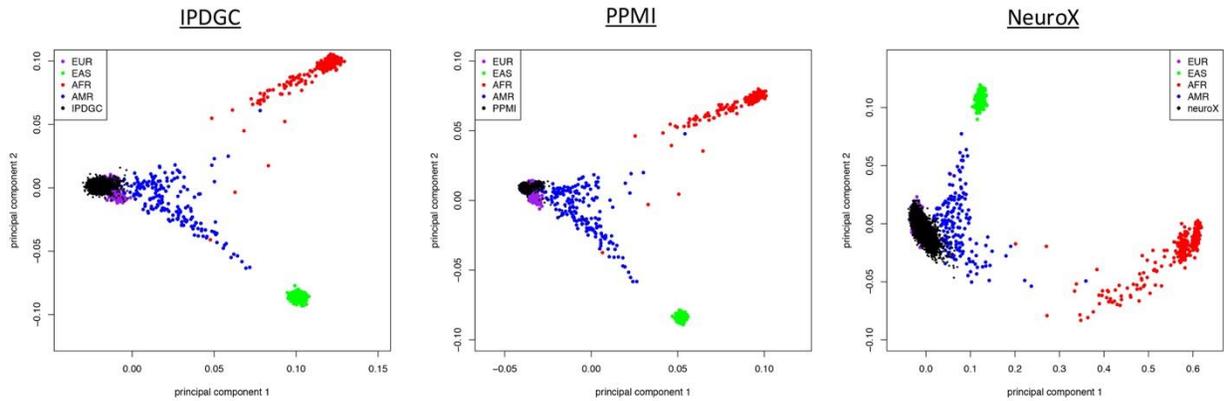
## SUPPLEMENTAL INFORMATION

### **Excessive burden of lysosomal storage disorder gene variants in Parkinson’s disease**

Laurie A. Robak, MD, PhD, Iris E. Jansen, PhD, Jeroen van Rooij, BSc, André G. Uitterlinden, PhD, Robert Kraaij, PhD, Joseph Jankovic, MD, International Parkinson’s Disease Genomics Consortium (IPDGC), Peter Heutink, PhD, Joshua M. Shulman, MD, PhD

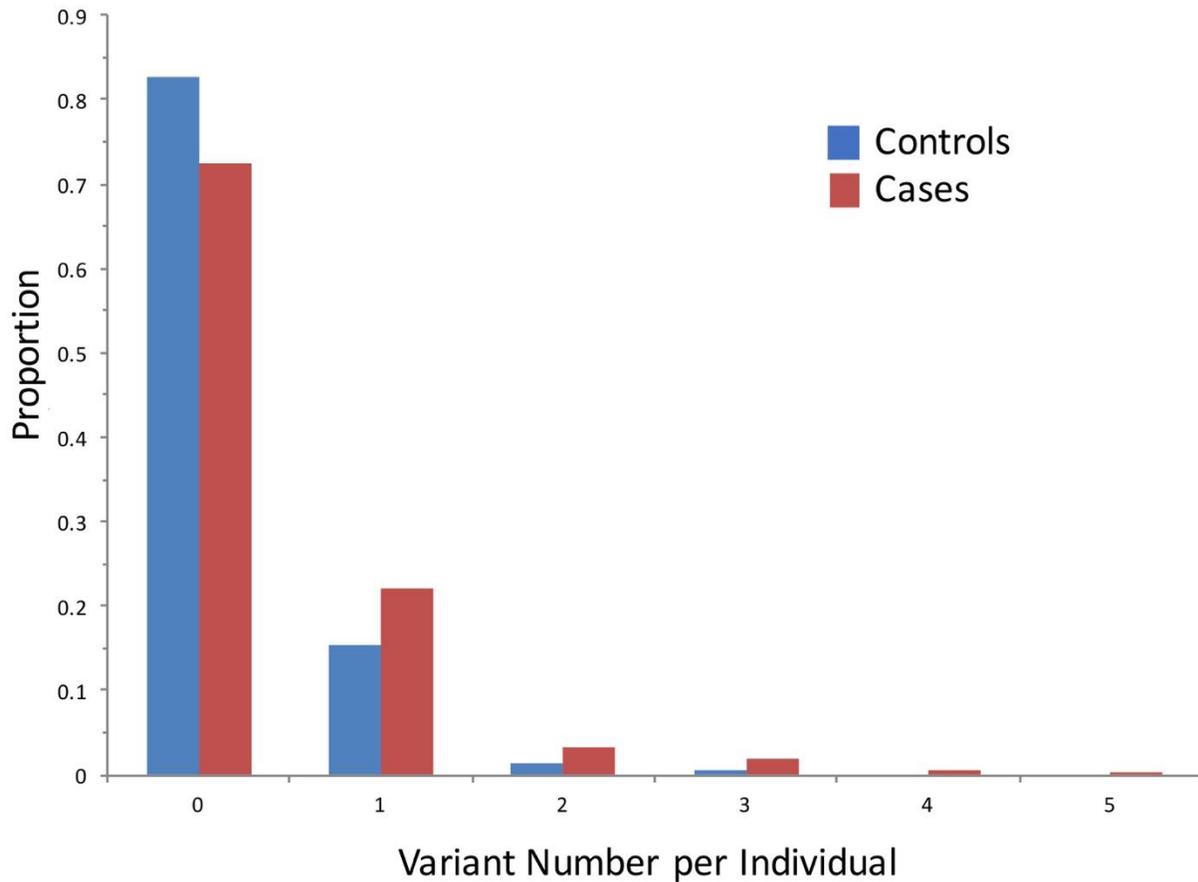
Supplementary Figure 1: Population Structure Analysis.....	2
Supplementary Figure 2: Distribution of variants within LSD gene drivers in the IPDGC cohort.....	3
Supplementary Table 1: Cohort clinical and demographic data.....	4
Supplementary Table 2: LSD Genes and Variant numbers in the Study Cohorts.....	5
Supplementary Table 3: LSD Gene Variants in the IPDGC Dataset .....	EXCEL file
Supplementary Table 4: Concordance rates for individuals with WES and genotype data.....	6
Supplementary Table 5: Unadjusted p-values for LSD Gene set Analyses.....	6
Supplementary Table 6: IPDGC Homozygous and Hemizygous Variants.....	7

## Supplemental Figure 1 Population Structure Analysis.



Multi-dimensional scaling component analysis based on linkage disequilibrium-pruned, genome-wide common variant markers is shown for each study cohort. Prior to these calculations, our datasets were merged with available genotypes from 1000 Genomes Project (1000GP) ancestry-based population samples, including African (AFR), East Asian (EAS), European (EUR) and the Americas (AMR) (1000 Genomes Project Consortium 2012). Following quality control filters, subjects in the discovery and replication samples (Black) are observed to cluster tightly with European ancestry subjects (Purple) on multi-dimensional scaling plots

**Supplemental Figure 2**  
**Distribution of variants within LSD gene drivers in the IPDGC cohort.**



The number of likely damaging variants ( $MAF < 3\%$ ,  $CADD\ C\text{-score} \geq 12.37$ ) per individual is shown versus the proportional representation in the IPDGC discovery cohort. Cases (Red) and Controls (Blue) are plotted separately. Many individuals harbor multiple alleles, and the distribution is right-skewed among Parkinson's disease cases. Compared to Figure 2, the analysis is restricted to those 5 genes prioritized as candidate drivers for the LSD gene set association with PD risk: *GBA*, *CTSD*, *SMPD1*, *SLC17A5*, and *ASAH1*.

**Supplementary Table 1 Cohort clinical and demographic data.**

	<b>cases</b>					<b>controls</b>		
	<i>n</i>	male	recruitment age (SD)	PD onset age (SD)	family history	<i>n</i>	male	recruitment age (SD)
<b>IPDGC</b>	1156	60.5%	51.5 (11.5) <sup>1</sup>	41.2 (10.8)	40.40%	1679	51.6%	63.7 (17.1)
<b>PPMI</b>	436	65.4%	61.7 (9.7)	59.8 (17.1)	27.10%	169	65.5%	61.8 (10.1)
<b>NeuroX</b>	6713	64.2%	NA	61.6 (12.4)	NA	5964	55.1%	64.1 (14.3)

<sup>1</sup>Recruitment age was available for 966 out of 1156 IPDGC Cases.

**Supplementary Table 2 LSD Genes and Variant numbers in the Study Cohorts.**

Disease	Gene	IPDGC			PPMI		NeuroX	
		Nonsyn	CADD	LoF	Nonsyn	CADD	Nonsyn	CADD
Aspartylglucosaminuria	<i>AGA</i>	13	10	2	4	3	7	6
Metachromatic Leukodystrophy	<i>ARSA</i>	5	5	1	3	2	5	5
Maroteaux-Lamy disease	<i>ARSB</i>	11	10	0	8	6	11	9
Farber Lipogranulomatosis	<i>ASAH1</i>	20	17	2	10	10	10	8
Kufor-Rakeb syndrome	<i>ATP13A2</i>	24	18	1	13	7	24	16
Neuronal Ceroid Lipofuscinosis (CLN3)	<i>CLN3</i>	18	17	3	4	4	5	4
Neuronal Ceroid Lipofuscinosis (CLN5)	<i>CLN5</i>	0	0	0	0	0	6	5
Neuronal Ceroid Lipofuscinosis (CLN6)	<i>CLN6</i>	10	7	1	6	5	10	9
Neuronal Ceroid Lipofuscinosis (CLN8)	<i>CLN8</i>	9	4	0	0	0	10	5
Cystinosis	<i>CTNS</i>	13	12	0	5	3	8	5
Galactosialidosis	<i>CTSA</i>	14	11	0	3	1	4	2
Neuronal Ceroid Lipofuscinosis (CLN10)	<i>CTSD</i>	7	4	0	2	1	5	3
Neuronal Ceroid Lipofuscinosis (CLN13)	<i>CTSF</i>	11	9	0	4	3	3	3
Pycnodysostosis	<i>CTSK</i>	6	5	0	3	2	3	3
Neuronal Ceroid Lipofuscinosis (CLN4B)	<i>DNAJC5</i>	5	5	1	0	0	1	1
Fucosidosis	<i>FUCA1</i>	15	12	0	7	5	5	4
Pompe disease	<i>GAA</i>	15	10	0	14	10	11	8
Krabbe disease	<i>GALC</i>	36	30	8	4	3	8	6
Morquio A disease	<i>GALNS</i>	22	14	0	8	2	10	6
Gaucher disease	<i>GBA</i>	39	32	4	8	7	11	9
Fabry disease	<i>GLA</i>	9	7	0	1	1	6	5
GM1-Gangliosidosis/Morquio B	<i>GLB1</i>	8	4	0	9	7	15	11
GM2-Gangliosidosis	<i>GM2A</i>	1	1	0	0	0	1	1
I-Cell disease	<i>GNPTAB</i>	39	31	13	10	8	15	11
Sanfilippo D syndrome	<i>GNS</i>	20	11	3	1	1	2	2
Neuronal Ceroid Lipofuscinosis (CLN11)	<i>GRN</i>	19	12	0	5	4	14	8
Sly disease	<i>GUSB</i>	17	10	0	5	3	3	2
Tay-Sachs disease	<i>HEXA</i>	20	18	1	4	3	12	11
Sandhoff disease	<i>HEXB</i>	8	6	2	2	1	10	7
Sanfilippo C syndrome	<i>HGSNAT</i>	18	15	1	7	6	7	4
Mucopolysaccharidosis Type IX	<i>HYAL1</i>	13	9	0	3	0	8	3
Hunter syndrome	<i>IDS</i>	9	8	2	0	0	8	5
Hurler syndrome	<i>IDUA</i>	8	4	0	2	1	12	11
Neuronal Ceroid Lipofuscinosis (CLN14)	<i>KCTD7</i>	4	3	0	1	1	3	2
Danon disease	<i>LAMP2</i>	9	7	1	1	1	4	3
Wolman disease	<i>LIPA</i>	14	10	2	3	2	3	2
Alpha-Mannosidosis	<i>MAN2B1</i>	12	11	2	14	10	22	17
Beta-Mannosidosis	<i>MANBA</i>	18	15	4	6	5	10	7
Mucopolidosis Type IV	<i>MCOLN1</i>	19	14	0	6	5	8	6
Neuronal Ceroid Lipofuscinosis (CLN7)	<i>MFSD8</i>	18	14	2	4	2	14	11
Schindler Disease/Kanzaki disease	<i>NAGA</i>	9	8	0	7	7	6	6
Sanfilippo B syndrome	<i>NAGLU</i>	10	9	3	2	1	5	3
Sialidosis	<i>NEU1</i>	0	0	0	3	2	5	5
Niemann-Pick Disease Type C1	<i>NPC1</i>	43	35	5	12	10	26	18
Niemann-Pick Disease Type C2	<i>NPC2</i>	2	2	0	3	2	6	5
Neuronal Ceroid Lipofuscinosis (CLN1)	<i>PPT1</i>	9	7	1	2	1	5	5
Sphingolipid-activator deficiency	<i>PSAP</i>	22	16	0	3	3	13	11
Action mycolonus-renal failure syndrome	<i>SCARB2</i>	10	7	0	3	0	10	7
Sanfilippo A syndrome	<i>SGSH</i>	10	8	0	4	3	10	7
Salla disease	<i>SLC17A5</i>	18	17	1	5	5	8	8
Niemann-Pick disease Type A/B	<i>SMPD1</i>	25	21	1	7	6	17	11
GM3-Gangliosidosis	<i>ST3GAL5</i>	11	11	1	2	2	3	2
Multiple Sulfatase Deficiency	<i>SUMF1</i>	0	0	0	2	2	3	2
Neuronal Ceroid Lipofuscinosis (CLN2)	<i>TPP1</i>	15	13	1	11	8	16	12

The number of variants (MAF < 3%) in each LSD gene is shown for the IPDGC discovery cohort and replication cohorts (PPMI and NeuroX), including total number of nonsynonymous variants (nonsyn) and likely damaging variants (CADD). For the IPDGC discovery dataset, the number of loss-of-function (LoF) variants is also indicated.

**Supplementary Table 4 Concordance rates for individuals with WES and genotype data.**

	<i>n</i> individuals	<i>n</i> overlapping variants		<i>n</i> non-missing genotype calls		concordance rates	
		all LSD genes	GBA	all LSD genes	GBA	all LSD genes	GBA
<b>IPDGC</b>	572	226	4	120155	2266	0.999983	1
<b>PPMI</b>	566	167	2	87749	1132	0.999966	1

Genotyping concordance rates based on 2 independent assays were computed for a subset of subjects in the IPDGC or PPMI WES cohorts for whom array-based genotyping data was also available. We considered concordance among all exonic variants detected among the LSD gene set or independently considering GBA.

**Supplementary Table 5 Unadjusted p-values for LSD gene set analyses.**

Cohort	# Cases	# Controls	Variants <sup>a</sup>	(a) MAF < 1%		(b) MAF < 3%	
				n <sup>b</sup>	<i>p</i> <sub>LSD</sub> ( <i>p</i> -GBA) <sup>c</sup>	n	<i>p</i> <sub>LSD</sub> ( <i>p</i> -GBA) <sup>c</sup>
<i>Discovery</i>							
<b>IPDGC</b>	1,167	1,685	(1) nonsyn	746 (709)	0.028 (0.197)	760 (721)	0.0035 (0.026)
			(2) CADD	585 (555)	0.022 (0.192)	596 (564)	0.0011 (0.0099)
			(3) LoF	69 (65)	0.464	- <sup>d</sup>	-
<i>Replication</i>							
<b>PPMI</b>	436	169	(1) nonsyn	243 (237)	0.024 (0.025)	256 (248)	0.16
			(2) CADD	179 (174)	0.098	187 (180)	0.281
<b>NeuroX</b>	6,713	5,964	(1) nonsyn	452 (443)	0.034 (0.027)	467 (456)	0.0001 (0.001)
			(2) CADD	338 (331)	0.057	348 (339)	0.0001 (0.020)

**Note:** Identical analyses are shown as in Table 2, except unadjusted p-values are reported. For all primary SKAT-O analyses with an unadjusted p-value < 0.05 for the LSD gene set, secondary analyses were performed excluding variants in GBA (p-GBA).

(a) Variants were classified into nested categories based on two frequency thresholds, MAF < 1% (a) or 3% (b), and three functional filters, all nonsynonymous (1), CADD likely damaging (2), and LoF (3).

(b) n = total number of LSD variant (number of variants excluding GBA). In parentheses, the number of variants excluding those in GBA are shown.

(c) Empirical SKAT-O p-values are based on 10,000 permutations following randomization of case/control status.

(d) No additional LoF variants met the relaxed frequency threshold (MAF < 3%).

LSD = lysosomal storage disorder; MAF = minor allele frequency; IPDGC = International Parkinson's Disease Genomics Consortium Discovery Cohort; PPMI = Parkinson's Progression Markers Initiative Replication Cohort; NeuroX = NeuroX exome array cohort; nonsyn = nonsynonymous variants; CADD = Combined Annotation Dependent Depletion; LoF = loss of function variants

**Supplementary Table 6 IPDGC homozygous and hemizygous variants.**

Gene	Position (chr:bp)	Alt	Ref	dbSNP	ExAC maf	Subjects (n)
CTSK	chr1:150771707	T	A	NA	NA	0
GBA	chr1:155206167	C	T	rs2230288	0.0098	1
GLB1	chr3:33110383	G	A	rs35289681	0.018	2
AGA	chr4:178360811	G	T	rs76491548	0.0096	0
HGSNAT	chr8:43053062	G	C	rs148632988	0.0066	1
GNPTAB	chr12:102190521	C	T	rs117566084	0.0038	0
GALC	chr14:88407876	A	T	NA	NA	0
HEXA	chr15:72638892	T	C	rs1800431	0.036	1
GLA	chrX:100653420	C	A	rs28935490	0.0034	5
GLA	chrX:100656740	C	T	rs104894845	0.00072	2
GLA	chrX:100658816	G	A	rs148158093	0.00025	1
LAMP2	chrX:119580269	A	C	rs141541387	0.0011	1
LAMP2	chrX:119581776	C	T	rs145169006	0.0012	5
LAMP2	chrX:119581851	T	A	rs138991195	0.00028	1
IDS	chrX:148571970	C	T	NA	0.000057	1
IDS	chrX:148585694	A	G	NA	NA	1

The number of subjects with homo- or hemi-zygous genotypes based on WES are shown.